ARL-CR-0803 ● JULY 2016

US Army Research Laboratory

Augmented REality Sandtable's (ARES's) Impact on Learning

by Tarah N Schmidt-Daly, Jennifer M Riley, Kelly S Hale,
David Yacht, and Jack Hart
Design Interactive, Inc.
3504 Lake Lynda Drive STE 400, Orlando, FL

under contract W911NF-14-C-0069

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer’s or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-CR-0803 ● JULY 2016

US Army Research Laboratory

Augmented REality Sandtable's (ARES's) Impact on Learning

by Tarah N Schmidt-Daly, Jennifer M Riley, Kelly S Hale,
David Yacht, and Jack Hart
Design Interactive, Inc.
3504 Lake Lynda Drive STE 400, Orlando, FL

under contract W911NF-14-C-0069

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)* July 2016	**2. REPORT TYPE** Final	**3. DATES COVERED (From - To)** September 2014–December 2015

4. TITLE AND SUBTITLE Augmented REality Sandtable's (ARES's) Impact on Learning	**5a. CONTRACT NUMBER** W911NF-14-C-0069
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Tarah N Schmidt-Daly, Jennifer M Riley, Kelly S Hale, David Yacht, and Jack Hart	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Design Interactive, Inc. 3405 Lake Lynda Drive STE 400, Orlando, FL 32817	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) US Army Research Laboratory ATTN: RDRL-HRT-A 2800 Powder Mill Road Adelphi, MD 20783-1138	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) ARL-CR-0803

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The use of augmented reality (AR) to supplement training tools, specifically sand tables, can produce highly effective systems at relatively low costs that can enhance operational capabilities. The interactive and adaptive nature of these technologies support distributed training for tactical engagement and enhanced-scenario customization. The Augmented REality Sandtable (ARES) is projected to enhance training and retention of spatial knowledge and spatial-reasoning skills over traditional maps (e.g., paper map and digitized 2-D displays) by providing multimodal and multisensory learning experiences. Empirical findings from an effectiveness evaluation indicate ARES supports significantly improved landmark-identification and distance-estimation performance as compared to a paper map and a 2-D digital display of a 3-D map. Additionally, users provided high ratings of perceived utility for ARES. This report details the methods used in the effectiveness evaluation, presents the research results, and provides a discussion of benefits of an AR-training solution as well as future research regarding AR and multimodal training-system design and development for the Military.

15. SUBJECT TERMS

ARES, augmented reality, sand table, map reading, terrain, visualization, topography

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Charles R Amburn
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	41	**19b. TELEPHONE NUMBER (Include area code)** 407-384-3901

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Acknowledgments

The authors would like to acknowledge the following people who were paramount in the coordination, planning, and study implementation of this project:

- MAJ Hector Barajas, MSG Kevin McSwain, and the cadre of the University of Central Florida Reserve Officers' Training Corps for providing support and study resources;
- CPT Jeffrey A Hansen, SFC Tarshekia McNear, and SFC Alsina Marangelli of the Headquarters and Headquarters Company, 143rd Sustainment Command (Expeditionary), for providing support and study resources;
- MAJ Jerry Mize and SFC John Hardwick of the US Army Research Laboratory Human Research and Engineering Directorate's Advanced Training and Simulation Division for their expertise, experience, and guidance during experimentation and protocol development; and
- Dr Jason D Moss, Mr Jeremy F Flynn, and Dr Michael W Boyce, who provided valuable input and guidance.

1. Introduction

The concept of a sand table, providing the capability to visualize and simulate the multiple options of a battlespace on a changeable surface, has a long history in battle planning (Smith 2009). As Smith (2009) suggests, low-resolution sand tables are still effective at modeling an analog of the relevant battlespace features such as personnel, terrain—and more importantly—the probability of decision outcomes. Today, sand-table exercises (STEXs) are critical for developing situational awareness of the battlefield as well as developing cognitive-reasoning skills about the dynamics of the battlefield elements (Kott et al. 2014). In addition, STEXs are cited as effective tactical-training tools with the potential to positively affect cognitive-skill development (e.g., spatial orientation) and tactical decision making (Amburn et al. 2015). Any effort made to enhance or modify existing, traditional sand tables would need to ensure the new system will meet ease-of-use and training outcome requirements deemed important by Military instructors and leaders.

There are high-tech alternatives to traditional sand tables that reflect the complexity and dynamically changing needs of current, diverse battlespaces. For example, multitouch surfaces (McManis 2012; Szymanski et al. 2008; Bortolaso et al. 2014) and total digitization concepts such as 3-D holographic displays (McIntire et al. 2014; Qi et al. 2005) purport to replace sand table capabilities in the field. These systems may augment sand table capabilities by providing consistent distributed training, improved after action reviews, and heightened environmental fidelity (Waller et al. 1998). These advancements do come at a substantially higher cost as system components (e.g., off-the-shelf multitouch surfaces and 3-D holographic displays) are available for thousands of dollars and must be integrated with customized software solutions. More importantly, these components come with scaling, brightness, resolution, and durability issues that may limit their use in the field (Geng 2013; Schoning et al. 2008). Beyond the technical issues, research is needed to understand the benefits and drawbacks in terms of effectiveness of training and efficiency of use of various technologies to execute STEXs and similar tactical exercises, with the goal of supporting acquisition cost-benefit analysis based on statistical outcomes regarding efficacy.

The Augmented REality Sandtable (ARES) system combines a traditional sand table, providing tactile and 3-D visualization of terrain, with a digital overlay that provides enriched graphics and interactivity (Amburn et al. 2015). This combination yields a multimodal and multisensory learning experience (O’Shea et al. 2009) over both paper maps and digital maps in isolation, which is expected to result in significantly higher spatial knowledge and spatial-reasoning skills acquisition.

This study was designed to empirically examine spatial-knowledge acquisition and understanding across 3 distinct media: 1) a 2-D paper map, 2) a digitized 2-D display of a 3-D map, and 3) a digital down-project 2-D map displayed on a 3-D sand table (ARES). Two main objectives of the study were to a) determine the impact of media tool on spatial-knowledge acquisition and spatial-reasoning skills and b) provide a foundation for establishment of a learner-centric environment using novel technology. Based on previous research, it was hypothesized that the 2-D map overlaid on a 3-D surface (ARES) would significantly improve spatial-knowledge acquisition over existent training mediums, as it provides additional visual-depth cues to enhance terrain visualization (Wickens et al. 2000) over map presentations alone. Improved spatial-knowledge acquisition was equated to improved performance scores across 3 experimental tasks—landmark identification, distance estimation, and situational judgements—which have been used extensively in the research to evaluate spatial knowledge (Darken and Peterson 2014; McDaniel and Whetzel 2007; Siegel and White 1975; Thorndyke and Goldin 1983).

2. Methods

2.1 Participants

Fifty participants—39 males and 11 females—ranging in age from 18 to 41 years (mean age [M_{age}] = 22.49, standard deviation of age [SD_{age}] = 4.35) voluntarily completed the study. Participants who were not active duty Military personnel were compensated for voluntary participation. Of the total participants, 43 were from the University of Central Florida's Reserve Officers' Training Corps and 7 were reserve personnel from the US Army's 143rd Sustainment Command in Orlando, Florida. All participants reported familiarity and experience with map reading. A within-subjects design was implemented in which all participants experienced the full range of experimental conditions. Informed consent was obtained from all participants.

2.2 Apparatus

Three experimental display media were used in this study:

1) A 2-D paper map that included an 8.5- × 11.0-inch representative portion taken from a US Army topographical map of Fort Irwin, California.
2) A 2-D digital 3-D map display that was a view of Google Earth (free desktop version) presented on a laptop computer showing an area of Yuma, Arizona.

3) A digital down-project 2-D map displayed on a 3-D sand table comprising an ARES system with a 55- × 32-inch sand table 3 ft high that contained sand 7.5 inches deep, and above it a projector down-projecting a 2-D display of an area of the Mustang Mountains near Fort Huachuca, Arizona.

The 3 test maps were developed by an Army subject matter expert (SME) with more than 25 years of service, including Air Defense operations, and were based on equivalency of terrain and landmarks present in the area. Expert agreement on equivalency of the terrain difficulty of the map areas and landmarks was established with a second Army SME with more than 20 years of service.

2.3 Tasks and Stimuli

For each experimental display medium, participants were asked to complete 3 tasks. The 3 tasks were completed in order as listed here for each participant on each medium to consistently present evaluations in order of spatial knowledge complexity (Darken and Peterson 2014).

2.3.1 Landmark Identification Test

First, the Landmark Identification Test consisted of 8 multiple choice items assessing an individual's proficiency in reading the various landmark features on a map that was based on work by Jones (1999). Participants answered on a tablet by selecting from a drop-down list of potential answers to define each specific landmark feature identified in the presented map.

2.3.2 Distance Estimation Test

Second, participants completed the Distance Estimation Test. This included 3 items assessing distance estimation (Darken and Banker 1998; Darken and Peterson 2014). Participants were asked to review 3 specific route segments marked on the provided map and estimate the driving distance, taking terrain into account. Each map included a scale for reference. Participants were asked to estimate the distances from a start location to Location 1, from Location 1 to Location 2, and from Location 2 to Location 3 (see Fig. 1). Participants provided their answers on a tablet by entering their response in a numeric field on the designated questionnaire form.

2.3.3 Situational Judgment Tests

Finally, each participant completed 2 Situational Judgment Tests (SJTs) designed to assess survey knowledge (Darken and Peterson 2014; McDaniel and Whetzel 2007). Each scripted scenario outlined a scenario objective and required participants to label the best, second best, and worst route alternatives displayed on the provided map given the scenario objectives and constraints. Each SJT was developed by the Army SME (described in Section 2.2).

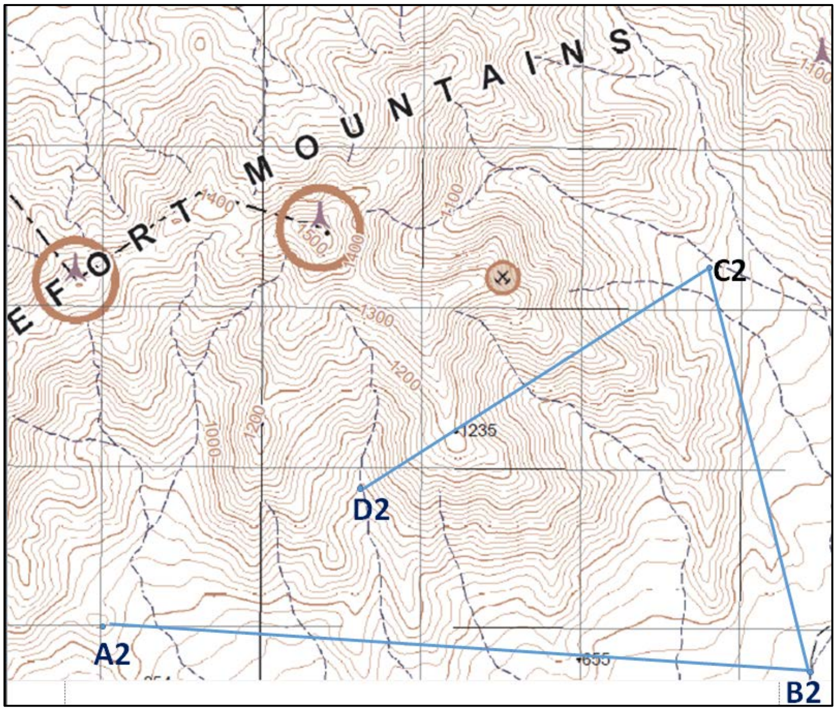

Fig. 1 Example of distance estimation task

2.4 Questionnaires, Surveys, Psychometric Tests, or Forms

Validated measures associated with individual differences were selected based on past research and theory to examine the effects of individual differences potentially impacting performance of the spatial knowledge tasks conducted in this study (Goldiez et al. 2007; Darken and Banker 1998; Hart and Staveland 1988; Watson et al. 1988). Demographics including age, gender, education, and experience were captured on a self-report questionnaire. Additional metrics included the following:

- Positive and Negative Affect Schedule (PANAS), a validated measure of mood during the past week, which may affect individuals' active engagement in each task (Watson et al. 1988).

- Task Self Efficacy, a self-report of one's perceived capability of his/her competency to perform spatial knowledge tasks using each tool, which was developed based on those used by Jones (1999).
- Spatial Orientation Aptitude, a validated, computerized assessment battery (Carpenter et al. 2010; Johnston et al. 2011) to evaluate the ability to imagine rearrangements or restructuring of individual components or surfaces to form a multidimensional object (Ekstrom et al. 1976). Past research has found that spatial orientation aptitude is related to higher learning and performance on the types of tasks in the present study (Goldiez et al. 2007; Darken and Banker 1998).
- NASA Task Load Index (NASA-TLX), a validated, perceived workload and effort self-report questionnaire to capture data against these variables, which could affect variance in learning and performance scores (Hart and Staveland 1988.)
- Utility Perceptions (UP) Questionnaire, a 5-item self-report that assess one's perceptions of the utility of each level of training and planning tool type for learning and potential for performance improvement (Davis 1989).

2.5 Experimental Design

The experiment was conducted as a within-subjects experimental design with a single factor consisting of 3 levels: paper map, 2-D display of 3-D map, and down project of 2-D on 3-D surface. Within the context of a series of spatial knowledge acquisition exercises, participant performance was assessed on multiple dependent measures: Landmark Identification Test, Distance Estimation Test, SJT, and UP (Darken and Banker 1998; Darken and Peterson 2014; Jones 1999; McDaniel and Whetzel 2007). Potential covariate measures collected included mood (Watson et al. 1988), task self-efficacy (adapted via Jones 1999), and spatial-orientation aptitude (Johnston et al. 2011). Measures of perceived workload (NASA-TLX) (Hart and Staveland 1988) and perceived utility (Davis 1989) were also collected.

2.6 Procedure

Upon arrival, participants reviewed and signed an informed-consent form. Participants were then provided with an overview of the study and asked to complete pretask and self-report measures, including the demographics questionnaire and the PANAS. After completing the self-report measures, individuals completed the computerized test, which assessed spatial orientation aptitude. During testing, study conditions were counterbalanced to avoid order effect. Each participant was exposed to the 3 study conditions: paper map, 2-D digital display of a 3-D map, and a 2-D overlay on a 3-D sand surface. Before completing tasks on each interface, participants completed the Task Self Efficacy questionnaire. Participants then performed each spatial-knowledge task (landmark identification, distance estimation, and situational judgment test) on each of the 3 study conditions. Following exposure to each condition, participants completed the NASA-TLX to assess perceived workload and effort for the given condition. After all tasks were completed under a given condition, participants also completed a utility perceptions assessment after which they had a 5-min rest period. A total of 3 experimenters collected data following a prescribed experimental protocol to avoid differences in experimental procedures across experimenters. Table 1 presents the ordered description of the experimental procedures experienced by each participant. Exposure to each tool type for in-test assessments was counterbalanced.

Table 1 Order of experimental procedures

Informed consent	
Study instructions	
Pretest assessments	Demographic Questionnaire PANAS Spatial Orientation Aptitude Test
In-test assessments (with counterbalanced exposure to tool type)	Task Self Efficacy Landmark Identification Test Distance Estimation Test SJT
Posttest assessments	NASA-TLX UP
Debrief	

3. Results and Discussion

3.1 Data Reduction and Analysis

To assess data for analyses, data were coded and scored as follows.

Landmark identification produced 8 responses for each participant that were coded numerically as correct (1) or incorrect (0), then summed to give a total score of 0–8. Higher scores indicate better performance.

Distance estimation produced raw distance estimates in meters for a path. Delta distance was calculated for the path by taking the absolute value of participant distance estimates from the actual distance values. The delta distance value indicates how close a participant’s estimates were to actual distances and, as such, smaller numbers indicate better distance-estimation performance. Three estimates were made along each and averaged to calculate delta distance scores. This approach was implemented based on distance-estimation tasks used in past research (e.g., Jones 1999) and provided opportunities to collect distance estimations across various terrain types (e.g., valleys, mountain range) within the smaller segments of the overall path. The goal was to provide multiple opportunities for estimating distances, increasing sampling, to better assess the impact of the tool type on the measure.

The SJT was scored by comparing participant rankings of route effectiveness to SME ranking of route effectiveness (see Table 2). Points were applied based on accuracy of participant ranking. When the optimal route per SME analysis of the criteria was ranked as optimal/first by a participant, 2 points were awarded. When the worst route was ranked as optimal/first by the participant, 0 points were awarded. If the second best route was ranked as optimal/first by the participant, 1 point was awarded.

Table 2 SJT scoring scheme

	SME-ranked routes	(A) Optimal	(B) Second best	(C) Worst
Participant-ranked route	**(a) Optimal**	2 points	1 point	0 points
	(b) Second best	1 point	2 points	1 point
	(c) Worst	0 points	1 point	2 points

Two situational judgment tasks were completed with a maximum of 6 points each, earned when participant rankings exactly matched SME rankings (see Table 3). Scores were summed across 2 situational judgment tasks as per guidance from past literature (McDaniel and Whetzel 2007), resulting in a total possible point value of 12. Higher scores indicate better performance.

Table 3 Scoring scheme

Participant-ranked route combinations	Points awarded
a/b/c	6
a/c/b	4
b/a/c	4
b/c/a	2
c/a/b	2
c/b/a	2

3.1.1 Outlier Analysis

Outliers were defined as data points that fell outside 2 standard deviations from the mean. If a single participant had more than 50% of his/her data considered to be an outlier, then the entire dataset for the participant would be omitted from data analysis (e.g., McGill et al. 1978). No participant was omitted completely by this rule. However, 5 outliers were detected for the distance estimation task and were excluded from that analysis only (see Appendix A). The goal for outlier removal was to reduce potential for inflated error rates, skewed data, and potential misrepresentation of statistical analysis (Zimmerman 1994).

3.1.2 Normality Testing

To test normality of the data, an evaluation of skewness and kurtosis was performed in the Statistical Package for Social Science. In addition, descriptive data were reviewed along with histograms with normalcy curves and box plots (see Appendix B). Variables that resulted in values of greater than ±2 in either skewness or kurtosis from the statistical evaluation were considered abnormal. (See Appendix C for a summary of the data.) According to the results, data from one variable, Utility Perception under the ARES condition, violated the criteria indicating skewness and kurtosis. The skewness of data is a result of very high ratings on subjective perceptions of utility for the ARES condition. Because there is no nonparametric equivalent, a repeated measures analysis of variance (ANOVA) was used to analyze the data.

3.1.3 Sphericity Testing

Additionally, Mauchly's sphericity test was performed for all repeated measures ANOVA tests performed, as sphericity is an important assumption for repeated measures ANOVA. The F-tests violation was found for Distance Estimation data only. A Greenhouse–Geisser correction was applied. No other violations to the test of sphericity were found. The Greenhouse–Geisser correction F-statistic and accompanying p-value for all other tests were equal to that of the statistics reported when sphericity was assumed based upon the total number of participants.

For each metric under evaluation, a repeated measures ANOVA was used to analyze the screened data. We followed a full counterbalanced design to control for order effects with repeated measures analysis. Data were collected on a total of 54 participants (groups of 9 for each condition). Our analysis was conducted on data for 50 participants due to incomplete data sets. Analysis revealed no significant correlations between order of exposure to the 3 conditions and any of the dependent variables, indicating no order effects.

Eta squared (η^2) and partial eta squared (η_p^2) are 2 of the commonly used measures of effect size in ANOVA, particularly for repeated measures designs (Lakens 2013; Bakeman 2005). When reporting effect sizes for ANOVAs, it is recommended to report partial eta squared, which may also be useful for comparing effects across different studies in contrast to eta squared (Lakens 2013). Partial eta squared is reported here as the appropriate effect-size statistic for repeated measures ANOVA with the rule of thumb for small, medium, and large effects, respectively, as 0.01, 0.06, and 0.14. As partial eta squared does not explain the size difference between each of the pairwise mean differences, we also include Cohen's (1988) effect-size values (*d*) with the convention for small, medium, and large effects, respectively, as 0.2, 0.5, and 0.8. Past research evaluating the effectiveness on instructional design features through meta-analysis of studies comparing different simulation-based instructional interventions utilized these Cohen's effect-size benchmarks for determining educational significance (Cook et al. 2013). Post hoc analysis was completed using Bonferroni's procedure and is reported on significant effects.

3.2 Results

3.2.1 Landmark Identification

Significant main effects were found with regard to tool type for performance on landmark identification: $F(2, 94) = 11.55$, $p < 0.001$, $\eta_p^2 = 0.197$ (see Fig. 2). Landmark identification was assessed via a count of correct identifications. Higher values indicate better performance scores in the landmark identification task. Post hoc analysis of scores on landmark identification using Bonferroni's procedure indicated that average performance using the 2-D map projected on the 3-D sand table ($M_{ST} = 5.04$, $SD_{ST} = 2.14$) was significantly better than performance using the paper map ($M_P = 4.25$, $SD_P = 1.85$, $p = 0.006$, 95% confidence interval (CI) [0.19, 1.40], $d = 0.39$) and significantly better than scores using Google Earth ($M_{GE} = 3.88$, $SD_{GE} = 1.10$, $p < 0.001$, 95% CI [0.48, 1.85], $d = 0.69$). The scores for landmark identification using paper map and Google Earth were not significantly different, $p = 0.30$.

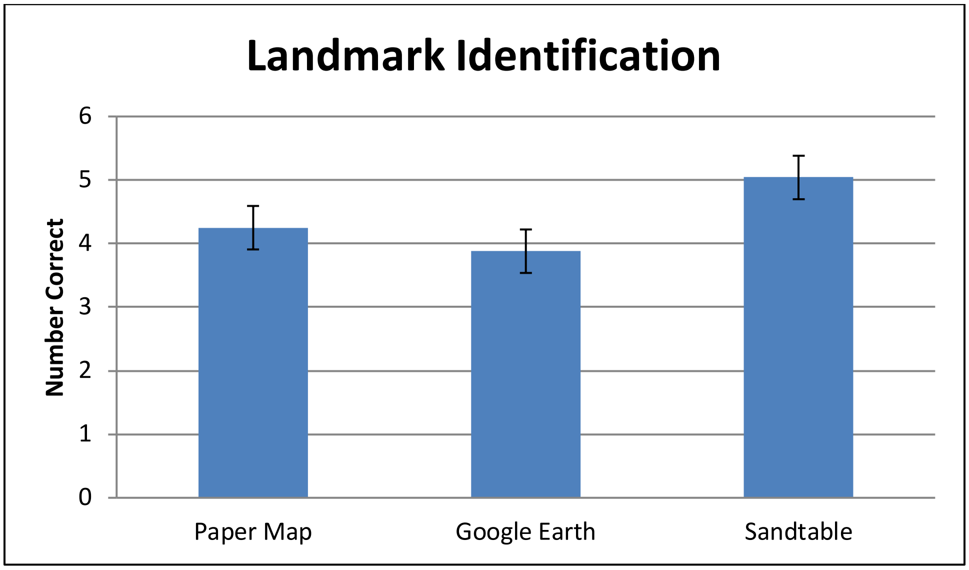

Fig. 2 Comparison of means for landmark identification performance (standard error [SE] bars shown)

3.2.2 Distance Estimation

There was a significant main effect of tool type for distance estimation performance (i.e., the accuracy in estimating distances as measured by the average deviation between participant estimates and actual distances), $F(1.71, 60.05) = 13.19$, $p < 0.001$, $\eta_p^2 = 0.274$ (see Fig. 3). Lower values indicate better performance in estimating distances across the tool types. Post hoc analysis of means using Bonferroni's procedure indicates that distance estimates performed using the 2-D map on the 3-D sand table ($M_{ST} = 296.27$, $SD_{ST} = 171.68$) were shown to be significantly better (more accurate) than paper ($M_P = 642.51$, $SD_P = 306.25$, $p < 0.001$, 95% CI [–486.44, –206.04], d =1.39) and Google Earth ($M_{GE} = 570$, $SD_{GE} = 433$, $p = 0.002$, 95% CI [–459.33, –88.11], $d = 0.83$). The paper map and Google Earth were not significantly different in terms of effects on distance estimation, $p = 1.00$.

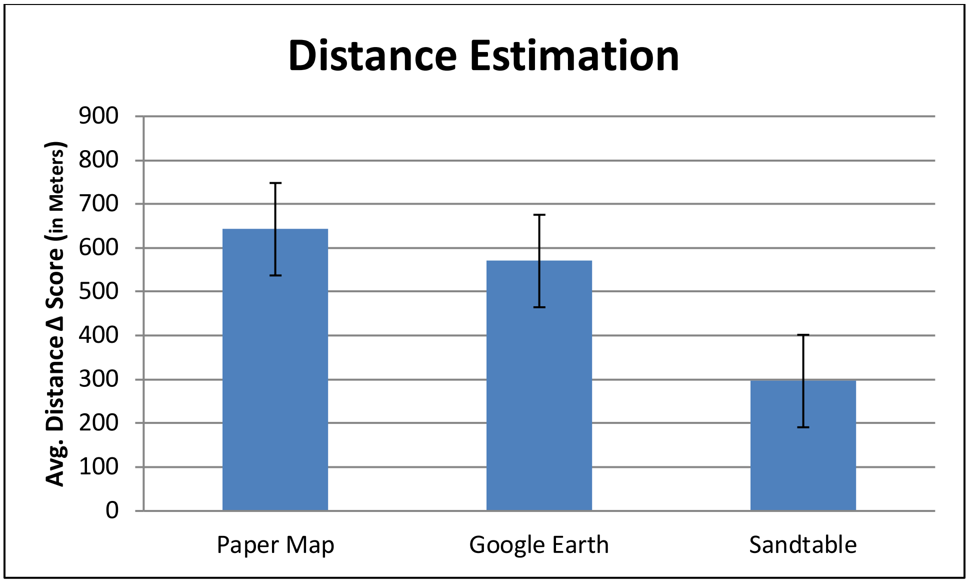

Fig. 3 Comparison of means for distance estimation performance (SE bars shown)

3.2.3 SJT

Statistical analysis on composite data across 2 SJTs revealed there was no significant effect of tool type on performance in SJTs, $F(2, 100) = 1.19$, $p = 0.309$, $\eta_p^2 = 0.023$. Performance with the sand table was comparable to performance using the other 2 media.

3.2.4 Perceived Utility

There was a significant main effect of tool type for UP, $F(2, 80) = 13.26$, $p < 0.001$, $\eta_p^2 = 0.249$ (Fig. 4). According to results of post hoc analysis, participants rated the utility of the sand table ($M_{ST} = 24.34$, $SD_{ST} = 4.99$) significantly higher as compared to either the paper map ($M_P = 19.10$, $SD_P = 6.68$, $p < 0.001$, 95% CI [2.33, 8.16], $d = 0.89$) or Google Earth ($M_{GE} = 19.44$, $SD_{GE} = 6.54$, $p < 0.001$, 95% CI [2.39, 7.42], $d = 0.84$). Perception of utility on the paper map and Google Earth were not significantly different, $p = 1.00$.

In general, participants rated the 2-D map projected on the 3-D sand table as highly useful. Specific user comments indicate the sand table was easy to use, that users enjoyed using the sand table to complete navigational tasks over traditional methods, and they thought they would learn wayfinding/navigational skills effectively using the sand table. Table 4 is a sample of participants' comments when asked open-ended questions about sand-table interaction.

Table 4 Subjective perceptions of sand table

What did you like best about the sand table?
Extremely easy route viewing, comparison of routes are much easier.
Fun, better than just looking at a map. Easy to use and see.
It showed in 3D what the contours looked like.
Seeing topographic features and elevation was much easier than on a paper map.
Super easy to learn on and gives you a good idea of what you should be doing in your head when you are navigating.
Better visualization.
Gives a more accurate representation of terrain than a map.
What did you like the least about the sand table?
Could be larger to incorporate more terrain features.
Just the fact that it has to be put in slideshow mode in order for it to project correctly.
It probably took a lot of time to form the terrain perfectly to the map.
Switching back between screens on the computer and the table.
What else would you like to tell us about the sand table as it relates to wayfinding/navigation?
Is a great tool and is very easy to use for navigating.
Good interactivity.
Very good learning tool to introduce people to a topographical map.
It is a good tool, especially for doing post-training after action reports.
That was AWESOME!!!! Where was this when I was in basic???? Especially helpful for people like me who struggle with depth perception.
Easily seen/user friendly.

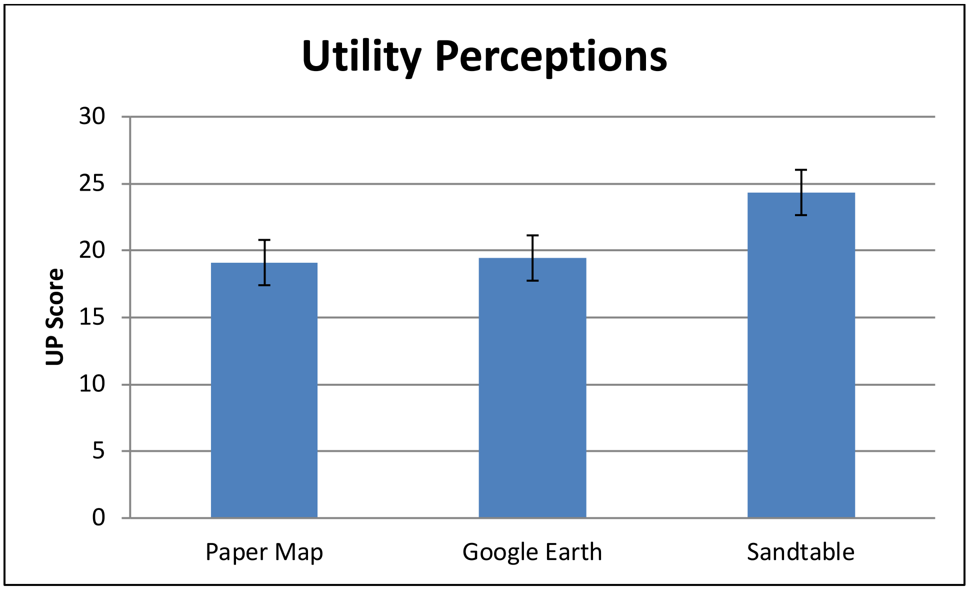

Fig. 4 Comparison of means for UP (SE bars shown)

3.2.5 Workload

The 2-D map projected on a 3-D sand table resulted in comparable subjective workload ratings (M_{ST} = 53.00, SD_{ST} = 13.16) as compared to traditional 2-D paper maps (M_P = 55.64, SD_P = 13.85) and Google Earth (M_{GE} = 54.91, SD_{GE} = 13.26). The sand table produced the lowest overall workload ratings, though the analysis of the workload data revealed no significant effect of tool type $F(2,96) = 1.23$, $p = 0.30$, $\eta_p^2 = 0.025$. This may be related to the perceived ease of use as captured in the UP questionnaire reported above.

4. Conclusions

The ARES combines the tactile nature of a traditional sand table with digital terrain overlay on sand to promote interactivity and improve terrain visualization in 3-D. This multimodal combination is expected to result in better performance on tasks involving spatial knowledge and spatial reasoning. The current study was aimed at assessing the impact of tool capabilities on these types of spatial skills and comparing ARES outcomes to performance results observed with more traditional map mediums: paper maps and 3D map representations (e.g., Virtual Battlespace). Results are promising and indicate that ARES, which provides for user-controlled, multimodal interactions with tactile and 3-D visualizations, produced superior performance on various spatial-knowledge tasks and higher overall perceived usefulness, on average, as compared to the other tool types. These results present initial evidence that ARES and similar novel technologies can be effective learner-centric environments for training and rehearsal.

In general, ARES was rated high in terms of perceived utility. Analysis revealed significantly higher ratings for ARES as compared to a paper map and 3-D representation. It is possible that the higher utility scores observed for ARES were the result of the novelty of the technology for participants. While this potential exists, it can be noted perceived utility was assessed for all media following counterbalanced exposure to the specific tool type and participants did not provide comparative assessments across the 3 tool types. Additionally, open-ended questions regarding the utility of ARES resulted in participant comments on the ease of use of the technology for completing the specified experimental tasks.

Though participants are expected to have been more familiar with paper maps, workload ratings for ARES were comparable to (not significantly higher than) ratings provided for the paper map and—also potentially novel—the 3-D map representation. This study found ARES produced significantly higher performance on landmark identification tasks and distance estimation tasks as compared to a paper map and a 3-D map representation. Surprisingly, similar results were not observed for the SJT. Analysis revealed no significant effect of tool type on these results. Upon further review of the test design and instructions, it was concluded that instructions for the situational judgment tasks may have produced a conflicting set of goals (i.e., prioritizing distance optimization with hazard avoidance) that the participants were not able to resolve in the manner expected. This limitation may have led to the overall low performance scores observed for the SJTs and, ultimately, the lack of impact of the tool type. Overall, ARES yielded superior results as compared to paper map and 3-D representation (e.g., Google Earth) on representative spatial knowledge and spatial reasoning tasks, suggesting it as a potential benefit to the US Military for training and planning.

Future research should focus on further exploring the findings presented here, specifically the inconclusive results on the situational judgment tasks and with respect to utility of ARES for enhancing performance on other spatial and decision making tasks (e.g., route planning and analysis, mission planning, course of action analysis), as well as tasks that incorporate interaction with the sand and digital overlay to assess the impact of the tangible interface to learning and knowledge retention. As interface designs can substantially impact the quality of performance with various technologies, work should be done to systematically evaluate ARES interfaces to ensure utility is not hindered by usability. Further, given the potential for ARES and other sophisticated sand-table technologies for enhancing training, research should be conducted to empirically evaluate a) the potential for learning, b) how long learning persists as compared to more traditional training tools, and c) the potential for transfer of training from ARES to reality.

5. References

Amburn CR, Vey NL, Boyce MW, Mize JR. The Augmented REality Sandtable (ARES). Orlando (FL): Army Research Laboratory (US); 2015 Oct. Report No.: ARL-SR-0340.

Bakeman R. Recommended effect size statistics for repeated measures designs. Behav Res Met. 2005;37(3):379–384.

Bortolaso C, Graham N, Scott SD, Oskamp M, Brown D, Porter L. Design of a multi-touch tabletop for simulation-based training. Proceedings of the 19th International Command and Control Research and Technology Symposium (ICCRTS), 2014 July; Alexandria (VA).

Carpenter A, Johnston M, Kokini C. CogGauge: A game-based cognitive assessment tool. International Conference on Human-Computer Interaction in Aerospace (HCI-Aero). 2010; Cape Canaveral (FL).

Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale (NJ): Lawrence Earlbaum Associates; 1988.

Cook DA, Hamstra SJ, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hatala, R. Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. Med Teach. 2013;35(1):e867–e898.

Darken RP, Banker WP. Navigating in natural environments: a virtual environment training transfer study. Proceedings of the IEEE 1998 Virtual Reality Annual International Symposium; 1998 March 14–18; Atlanta (GA). p. 12–19.

Darken RP, Peterson B. Spatial orientation, wayfinding, and representation. In: Hale KS, Stanney KM, editors. Handbook of virtual environments: design, implementation, and applications. 2nd ed. Boca Raton (FL): Taylor and Francis Group; 2015. p. 467–491.

Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quart. 1989;13(3):319–340.

Ekstrom RB, French JW, Harman HH, Dermen D. Manual for kit of factor-referenced cognitive tests. Princeton (NJ): Educational Testing Service; 1976.

Geng J. Three-dimensional display technologies. Adv Opt Phot. 2013;5(4):456–535.

Goldiez BF, Ahmad AM, Hancock PA. Effects of augmented reality display settings on human wayfinding performance. IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews. 2007;37(5):839–845.

Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N, editors. Human mental workload. Oxford, England: North-Holland; 1988. p. 139–183.

Johnston MR, Carpenter AC, Hale KS. Test-retest reliability of CogGauge: a cognitive empirical tool for spaceflight. Paper presented at: HCII Conference; 2011 July; Orlando (FL).

Jones QB. The transfer of spatial knowledge from virtual to natural environments as a factor of map representation and exposure duration [master's thesis]. [Monterey (CA)]: Naval Postgraduate School; 1999.

Kott A, Buchler N, Schaefer KE. Kinetic and cyber. In: Kott A, Wang C, Erbacher RF, editors. Cyber defense and situational awareness. New York (NY): Springer; 2014. p. 29–45)

Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Fron Psych. 2013;4(863):1–12.

McDaniel MA, Whetzel DL. Situational judgment tests. In: Whetzel DL, Wheaton GR, editors. Applied measurement: industrial psychology in human resources management. Mahwah (NJ): Lawrence Erlbaum and Associates; 2007. p. 235−257.

McGill R, Tukey JW, Larsen WA. Variations on box plots. Amer Stat. 1978;32:12–16.

McIntire JP, Havig PR, Geiselman EE. Stereoscopic 3D displays and human performance: a comprehensive review. Displays. 2014;35(1):18–26.

McManis LD. Sand and water table have a new friend: multi-touch table. Rosemount (MN): Dakota County Technical College; 2012 Nov 21 [accessed 2016 June 8]. http://blogs.dctc.edu/dawnbraa/2012/11/21/sand-and-water-table-have-a-new-friend-multi-touch-table/.

O'Shea P, Mitchell R, Johnston C, Dede C. Lessons learned about designing augmented realities. Int J Gam Comp Med Sim. 2009;1(1):1–15.

Qi W, Martens J, van Liere R, Kok A. Reach the virtual environment—3D tangible interaction with scientific data. Proceedings of the OZCHI; 2005 Nov; Canberra, Australia. p. 1–10.

Schoning J, Brandl P, Daiber F, Echtler F, Hilliges O, Hook J, Lochtefeld M, Motamedi N, Muller L, Olivier P, Roth T, von Zadow U. Multi-touch surfaces: a technical guide. Technical University of Munich (Germany); 2008. Report No.: TUM-10833.

Siegel AW, White SH. The development of spatial representation of large-scale environments. In: Reese HW, editor. Advances in child development and behavior. Vol. 10. New York (NY): Academic Press; 1975. p. 9–55.

Smith R. The long history of gaming in military training. Sim Gam. 2009;41(1):6–19.

Szymanski R, Goldin M, Palmer N, Beckinger R, Gilday J, Chase T. Command and control in a multitouch environment. Proceedings of the Army Science Conference; 2008 Dec; Orlando, FL.

Thorndyke PW, Goldin SE. Spatial learning and reasoning skill. In: Pick HL Jr, Acedolo CP, editors. Spatial orientation. New York (NY): Plenum; 1983. p. 195–217.

Waller D, Hunt E, Knapp D. The transfer of spatial knowledge in virtual environment training. Presence. 1998;7(2):129–143.

Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psych. 1988;54:1063–1070.

Wickens CD, Thomas LC, Young R. Frames of reference for the display of battlefield information: judgment-display dependencies. Hum Fac. 2000;42(4):660–675.

Zimmerman DW. A note on the influence of outliers on parametric and nonparametric tests. J Gen Psych. 1994;12(4):391–401.

INTENTIONALLY LEFT BLANK.

Appendix A: Graphs of Outlier Analysis

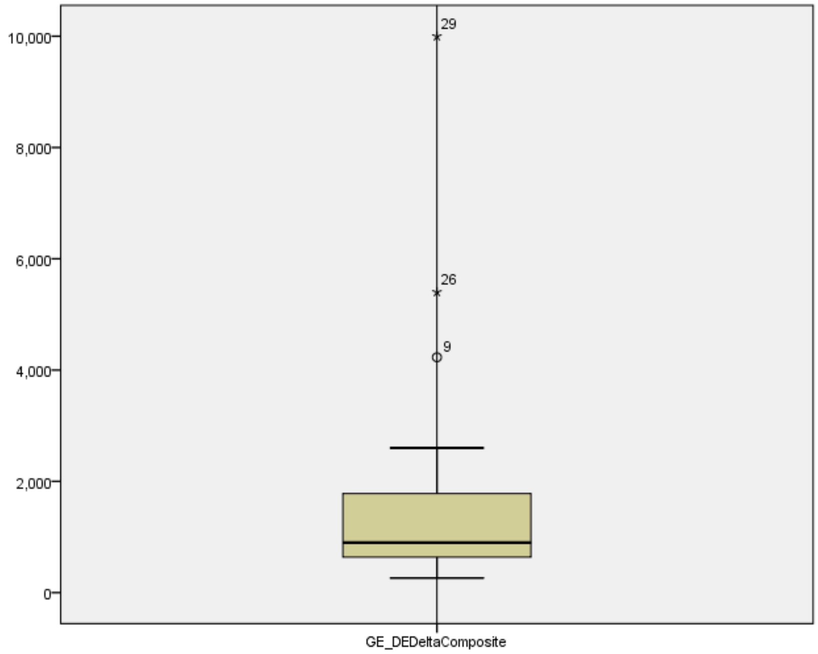

Fig. A-1 Outliers distance estimation scores, Google Earth condition; Participants 29, 26, and 9

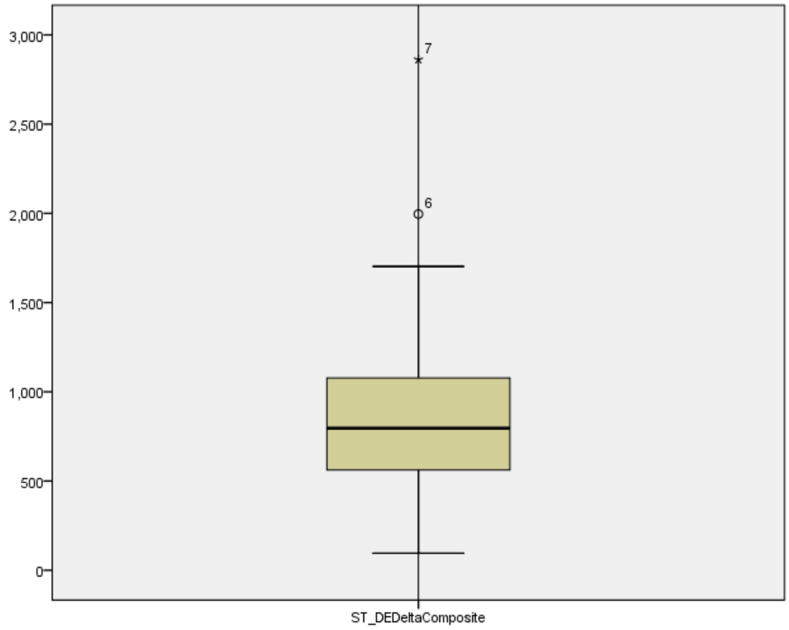

Fig. A-2 Outliers distance estimation scores, ARES sand-table condition; Participants 6 and 7

Appendix B: Histograms with Normalcy Curves

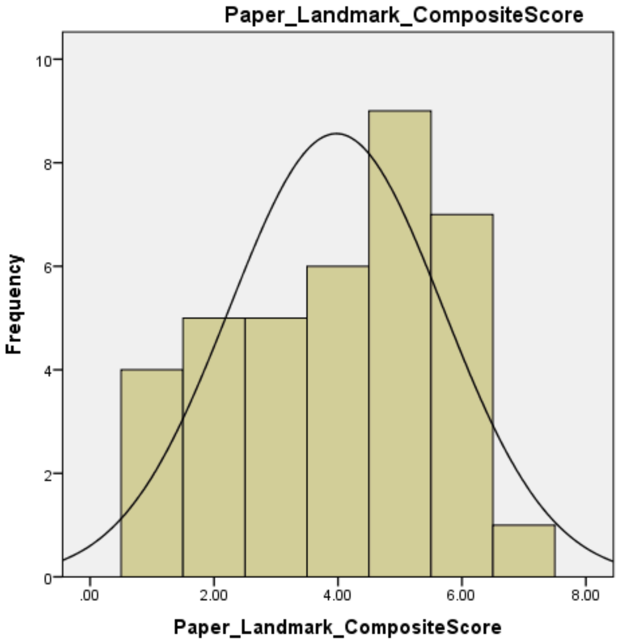

Fig. B-1 Histogram landmark identification scores, paper condition

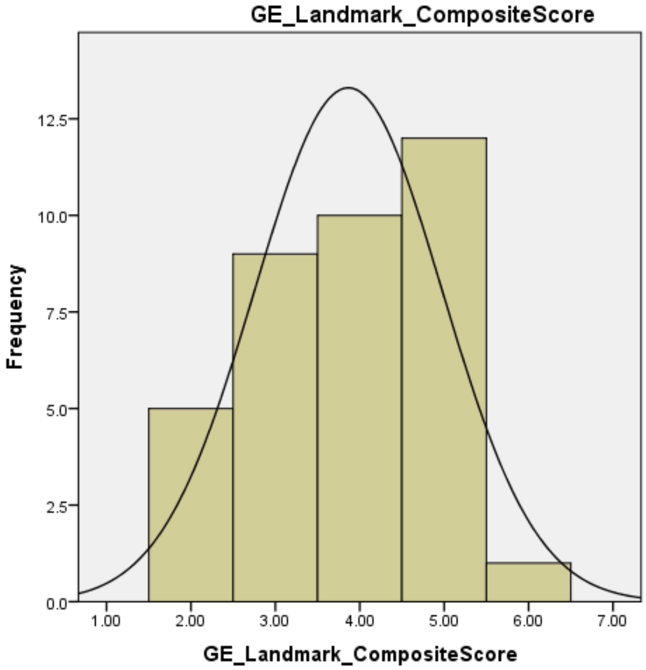

Fig. B-2 Histogram landmark identification scores, Google Earth condition

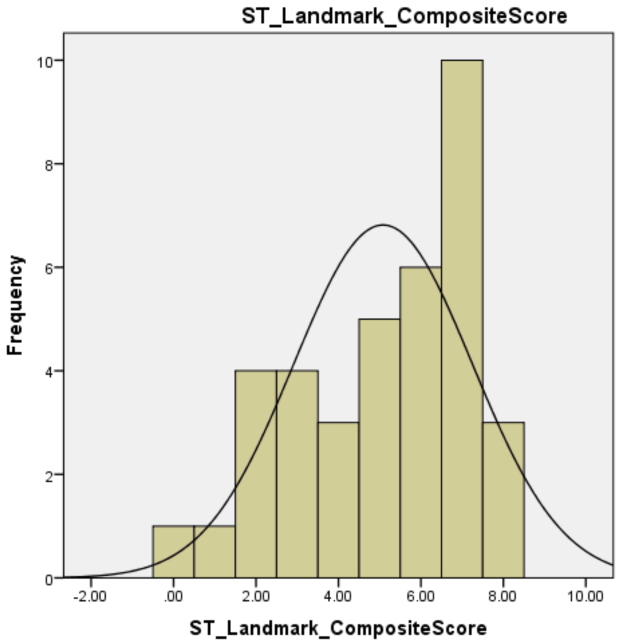

Fig. B-3 Histogram landmark identification scores, sand-table condition

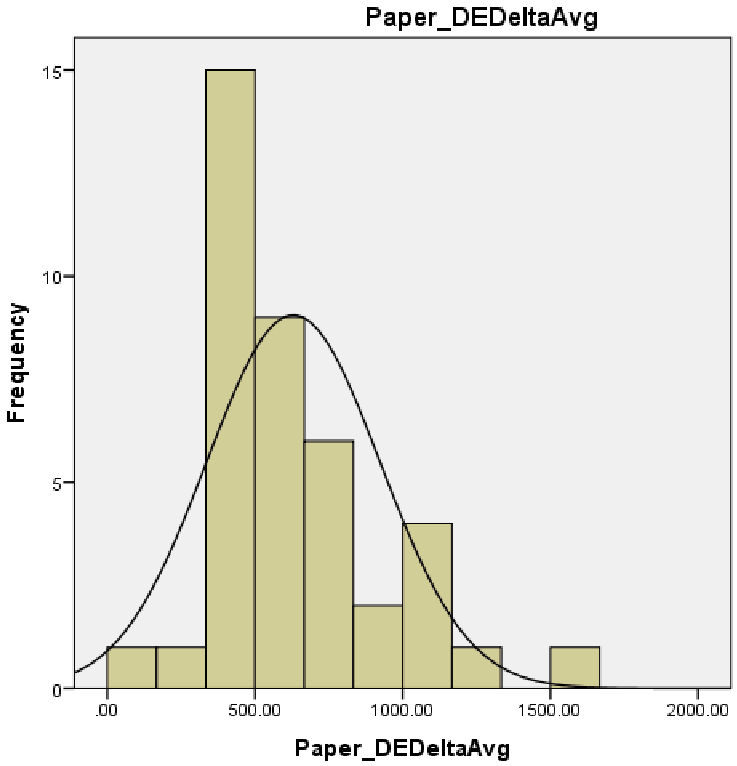

Fig. B-4 Histogram distance estimation scores, paper condition

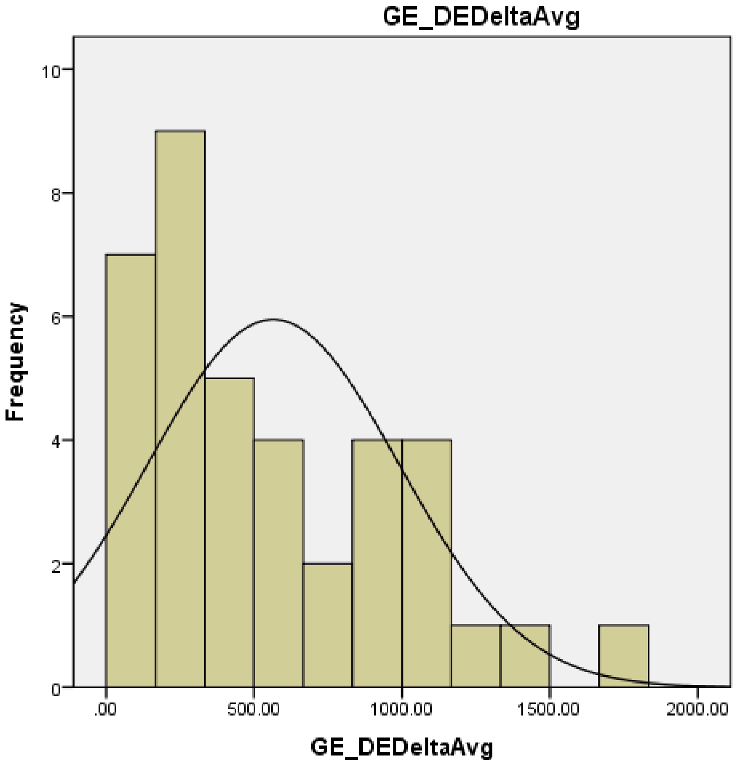

Fig. B-5   Histogram distance estimation scores, Google Earth condition

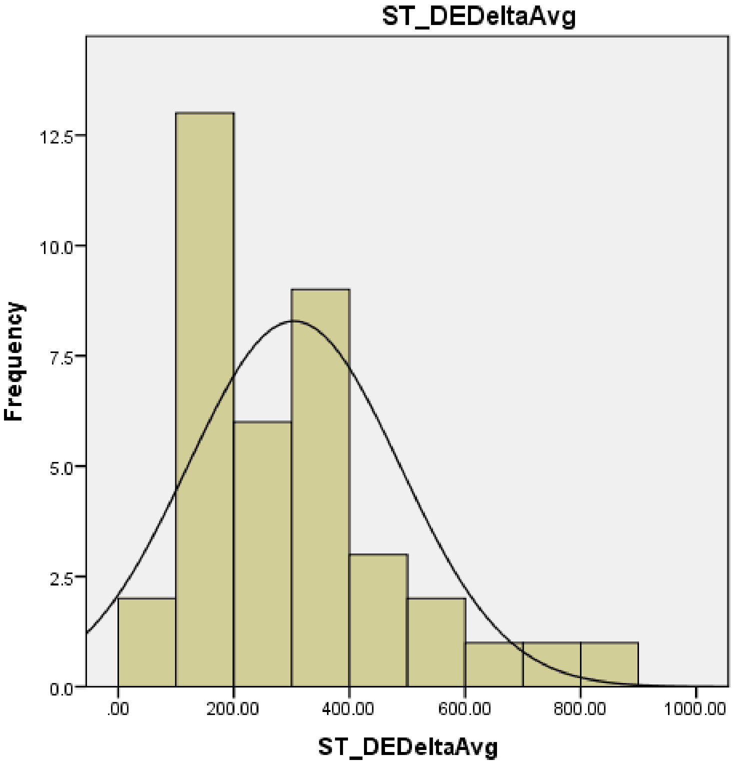

Fig. B-6 Histogram distance estimation scores, sand-table condition

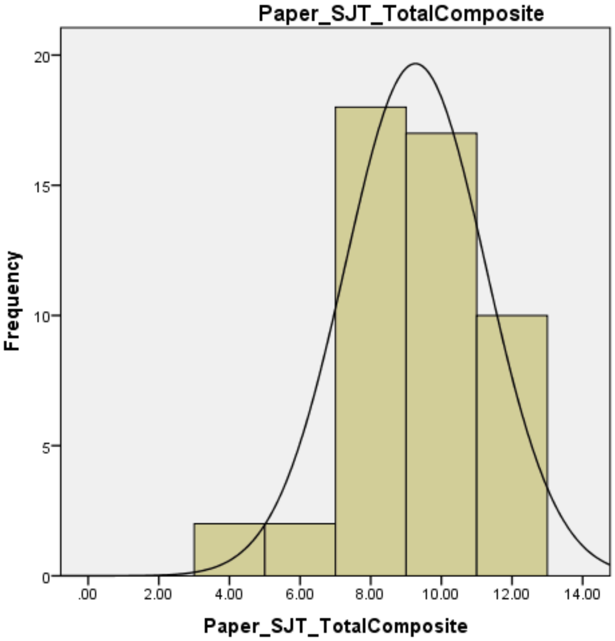

Fig. B-7 Histogram situational judgment task scores, paper condition

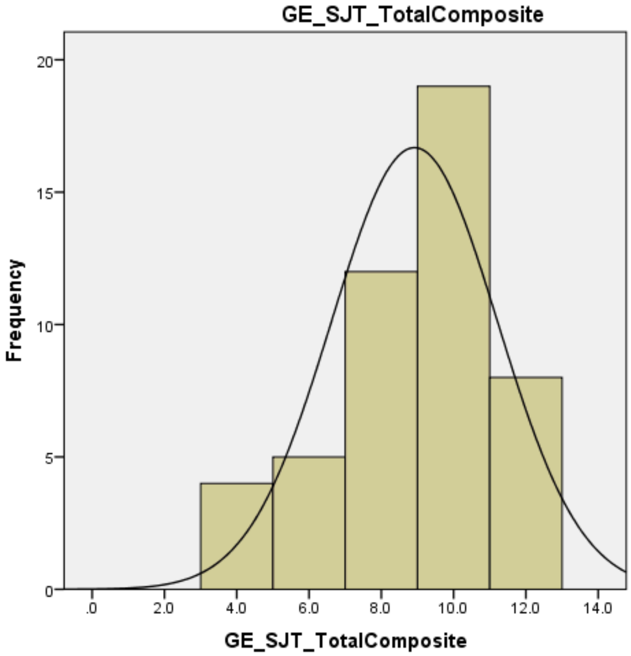

Fig. B-8 Histogram situational judgment task scores, Google Earth condition

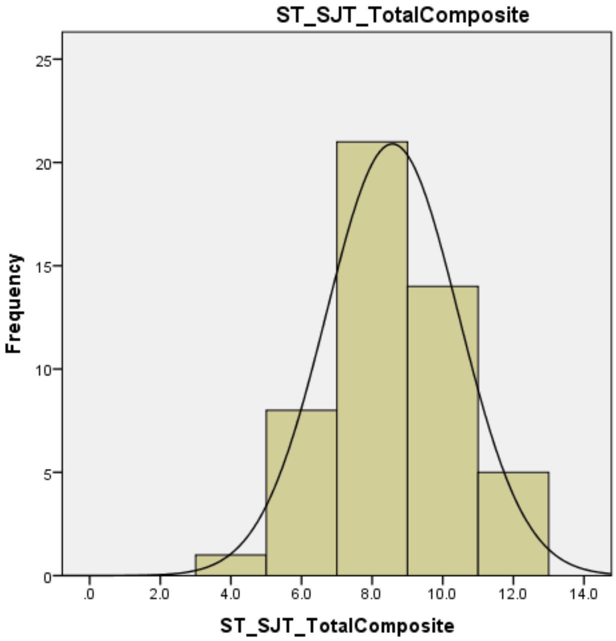

Fig. B-9 Histogram situational judgment task scores, ARES condition

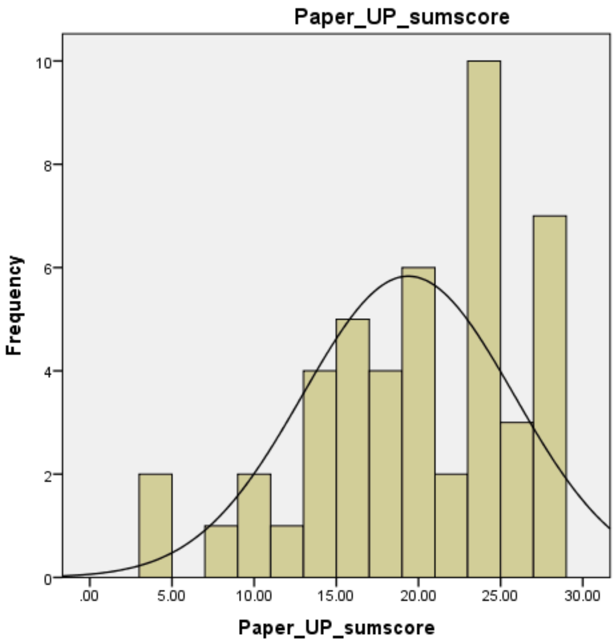

Fig. B-10 Histogram utility perception ratings, paper condition

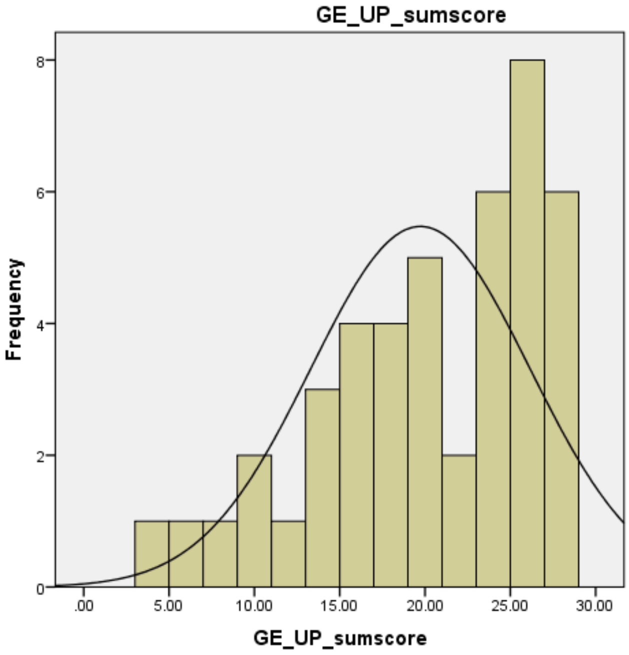

Fig. B-11 Histogram utility perception ratings, Google Earth condition

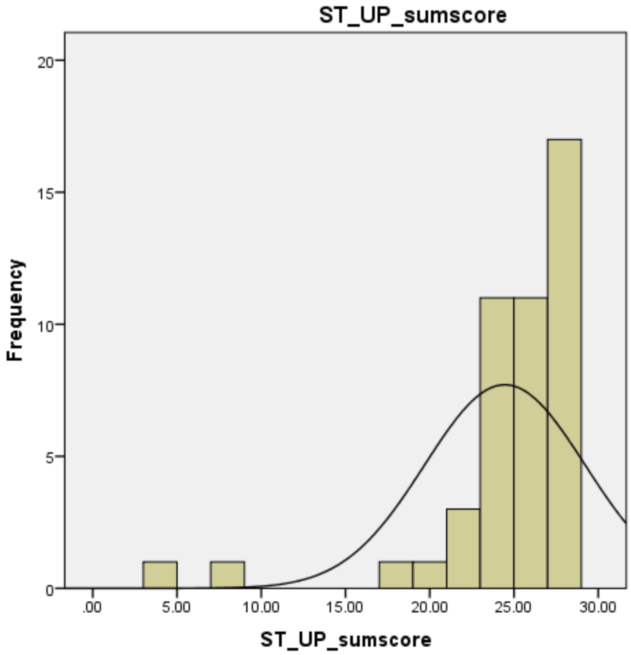

Fig. B-12 Histogram utility perception ratings, sand-table condition

Appendix C: Summary Results of Normality Testing

This appendix appears in its original form, without editorial change.

Table C-1 Normality test statistics

		Paper_Landmark_CompositeSCORE	Paper_DEDeltaAvg	Paper_SJT_TotalComposite	Paper_UP_sumscore	GE_Landmark_CompositeSCORE	GE_DEDeltaAvg	GE_SJT_TotalComposite	GE_UP_sumscore	ST_Landmark_CompositeSCORE	ST_DEDeltaAvg	ST_SJT_TotalComposite	ST_UP_sumscore
N	Valid	44	40	44	42	43	38	43	39	44	38	44	41
	Missing	13	17	13	15	14	19	14	18	13	19	13	16
Mean		4.3636	629.5833	9.2727	19.2143	3.8605	564.5263	8.837	19.8974	5.0682	303.9912	8.636	24.3171
Median		5.0000	539.3333	10.0000	20.0000	4.0000	461.1667	10.000	22.0000	6.0000	278.3333	8.000	26.0000
Mode		5.00	362.33[a]	8.00	23.00[a]	4.00	212.67[a]	10.0	25.00	6.00[a]	154.00	8.0	28.00
Std. Deviation		1.93007	293.81123	1.83460	6.52775	1.12507	424.59499	2.3997	6.67988	2.15015	183.02614	1.8689	5.00719
Variance		3.725	86325.041	3.366	42.611	1.266	180280.905	5.759	44.621	4.623	33498.567	3.493	25.072
Skewness		-.586	1.283	-.329	-.640	-.555	.824	-.536	-.714	-.651	1.119	.022	-2.652
Std. Error of Skewness		.357	.374	.357	.365	.361	.383	.361	.378	.357	.383	.357	.369
Kurtosis		-.628	1.989	.305	-.221	-.242	-.050	-.454	-.444	-.667	1.047	-.072	8.207
Std. Error of Kurtosis		.702	.733	.702	.717	.709	.750	.709	.741	.702	.750	.702	.724
Percentiles	25	3.0000	420.0833	8.0000	15.0000	3.0000	209.6667	8.000	15.0000	3.0000	158.0000	8.000	23.0000
	50	5.0000	539.3333	10.0000	20.0000	4.0000	461.1667	10.000	22.0000	6.0000	278.3333	8.000	26.0000
	75	6.0000	728.6667	10.0000	24.0000	5.0000	929.2500	10.000	25.0000	7.0000	398.6667	10.000	28.0000

a. Multiple modes exist. The smallest value is shown

List of Symbols, Abbreviations, and Acronyms

2-dimensional	2-D
3-dimensional	3-D
ANOVA	analysis of variance
ARES	Augmented REality Sandtable
CI	confidence interval
d	Cohen’s d
η^2	eta squared
$\eta_p{}^2$	partial eta squared
M	mean
NASA-TLX	NASA Task Load Index
PANAS	Positive and Negative Affect Schedule
SD	standard deviation
SE	standard error
SJT	Situational Judgment Test
SME	subject matter expert
STEX	sand-table exercise
UP	utility perceptions

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

2 DIRECTOR
(PDF) US ARMY RSRCH LAB
RDRL CIO LL
IMAL HRA MAIL & RECORDS MGMT

1 GOVT PRINTG OFC
(PDF) A MALHOTRA

1 SR INST SPEC
(PDF) US ARMY RSRCH LAB
RDRL HRT A
C AMBURN

INTENTIONALLY LEFT BLANK.

Approved for public release; distribution unlimited.